

13 Benford's Law

In 1881, Simon Newcomb had noticed that in tables of logarithms, the first pages were much more worn and smudged than later pages. In 1938, Frank Benford published a paper showing the distribution of the leading digit in many disparate sources of data. In all these sets of data, the number 1 was the leading digit about 30% of the time.

Benford's law has been found to apply to population numbers, death rates, lengths of rivers, mathematical distributions given by some power law, and physical constants like atomic weights and specific heats. This law is now used to detect fraud in lists of socio-economic data submitted in support of public planning decisions and as an indicator of accounting and expenses fraud.

In this programming assignment you will verify Benford's law for the US Census data of 2009. The file [Census 2009](#) gives the population distribution in the US. Each line of data has the name of the state, the town or village, and its population.

You will create a dictionary that create a frequency distribution of the the first digit of the population numbers. You will print out the actual frequency and the relative frequency of each digit. The sample output will look like:

Digit	Count	%
1	18	30.0
2	8	13.3
3	8	13.3
4	6	10.0
5	10	16.7
6	5	8.3
7	2	3.3
8	1	1.7
9	2	3.3